

<https://helda.helsinki.fi>

Waveform prototype-based feature learning for automatic detection of the early repolarization pattern in ECG signals

Tobon-Cardona, Marcela

2018-12

Tobon-Cardona , M , Kenttä , T , Porthan , K , Tikkanen , J T , Oikarinen , L , Viitasalo , M , Salomaa , V , Huikuri , H , Junttila , J M & Seppänen , T 2018 , ' Waveform prototype-based feature learning for automatic detection of the early repolarization pattern in ECG signals ' , Physiological Measurement , vol. 39 , no. 11 , 115010 . <https://doi.org/10.1088/1361-6579/aaecef>

<http://hdl.handle.net/10138/311998>

<https://doi.org/10.1088/1361-6579/aaecef>

unspecified

acceptedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

Waveform prototype-based feature learning for automatic detection of the early repolarization pattern in ECG signals

Marcela Tobón-Cardona¹, Tuomas Kenttä², Kimmo Porthan³, Jani T. Tikkanen², Lasse Oikarinen³, Matti Viitasalo³, Veikko Salomaa⁴, Heikki V. Huikuri², Juhani M. Junttila² and Tapio Seppänen¹

¹ Center for Machine Vision and Signal Analysis, University of Oulu, Oulu, Finland

² Research Unit of Internal Medicine, Medical Research Center Oulu, University of Oulu and Oulu University Hospital, Oulu, Finland

³ Division of Cardiology, Heart and Lung Center, Helsinki University Central Hospital, Helsinki, Finland

⁴ National Institute for Health and Welfare, Helsinki, Finland

E-mail: marcela.tobon@oulu.fi

Abstract

Objective: Our aim was to develop an automated detection method, for prescreening purposes, of early repolarization (ER) pattern with slur/notch configuration in electrocardiogram (ECG) signals using a waveform prototype-based feature vector for supervised classification.

Approach: The feature vectors consist of fragments of the ECG signal where the ER pattern is located, instead of abstract descriptive variables of ECG waveforms. The tested classifiers included linear discriminant analysis (LDA), k-nearest neighbor (KNN) algorithm, and support vector machine (SVM). *Main results:* SVM showed the best performance in Friedman tests in our test data including 5676 subjects representing 45408 leads. Accuracies of the different classifiers showed results well over 90%, indicating that the waveform prototype-based feature vector is an effective representation of the differences between ECG signals with and without the ER pattern. Accuracy of inferior ER was 92.74% and 92.21% for lateral ER. Sensitivity achieved was 91.80% and specificity was 92.73%. *Significance:* The algorithm presented here showed good performance results indicating that it could be used as a prescreening tool of ER, and it provides an additional identification of critical cases based on the distances to the classifier decision boundary, which are close to the 0.1 mV threshold and are difficult to label.

Keywords: early repolarization, ECG, waveform prototype, supervised learning

1. Introduction

Early repolarization (ER) pattern was considered a benign finding until 2008, when it was associated with Sudden Cardiac Arrest (SCA), calling the attention of the medical community [1]. Recent studies have shown that the ER pattern

is a specific sign of increased vulnerability to ventricular tachyarrhythmias [2] and risk for cardiac death and SCA, especially when accompanied by J point elevation in the inferior leads and notching configuration [3], [4].

In 2015, an agreed definition of the electrocardiographic ER pattern with notching and slurring configuration was published. According to the consensus definition, the notch/slur ER pattern is present if three conditions are met: i) The pattern lies above the baseline and there is an end-QRS notch or slur on the down-slope of a prominent R-wave, ii) J peak (J_p) is ≥ 0.1 mV in ≥ 2 contiguous leads of the 12-lead ECG, excluding leads V1 to V3, and iii) QRS duration is < 120 ms. [5]

Automatic methods to detect ER pattern have been proposed and are characterized by the analysis and quantification of the notch or slur slope [6], [7]. In contrast, electrocardiographic machines implemented commonly in clinical practice use interpretation algorithms to detect ST-segment elevation in addition to elevation of the J point [8], but do not include the ER notch/slur analysis.

Supervised classification has been used in medical applications with ECG signals, to assign each individual heartbeat to its specific class, or to detect a patient's pathology using the information contained in a set of features [9]. The features usually consist of parameters that can be measured from the signals (e.g. heart rate, frequency analysis, features derived from the shape of the ECG, etc.), information related to the patient (e.g. age, gender, weight, etc.) or combination of both. More sophisticated descriptors like wavelet-based features, parametric modelling approaches and features based on non-linear dynamics, have been proposed for ECG signals and are selected according to the application [9]. However, research in cognition has shown that humans categorize objects based on hierarchical comparison using a generic prototype [10]. In this way, prototype-based features can be used as a point of reference, to compare new data with the samples, and perform the classification.

Thus, we propose a waveform prototype-based feature vector to represent ER positive and negative cases. Using these prototypes, we have develop an automated detection algorithm of ER pattern with notch/slur configuration. Contrary to current methods, the prototype or feature vector is defined as a fragment of the ECG waveform, instead of the characterization of specific morphological details of the R-wave down-slope shape. The advantage of our approach is that no explicit feature extraction needs to be implemented as long as the proper fragment can be reliably found in the ECG. The fragment is located in the down-slope of R-wave, and effective algorithms have been published in literature [11] that can be used to detect this part of ECG robustly, if proper filtering is implemented first. We evaluate the classifier using a large data set, including 5676 subjects, and compare the performance with a state-of-the-art technique [7], putting special emphasis on sensitivity, as the method is proposed for prescreening purposes.

2. Methods

2.1 Data collection and manual grading of ER

Study subjects were drawn from the Health 2000 Survey, which is a general population-based survey conducted in Finland between 2000 and 2001. The study enrolled 8028 adults (≥ 30 yrs.) and was representative of the Finnish adult population at the time. Health examinations, including digital 12-lead ECG, were performed on 6354 subjects. The Institutional ethics committee of Helsinki and Uusimaa hospital district approved the study, which adhered to all principles outlined in the Declaration of Helsinki. Detailed description of the study protocol has been published in [12], [13]. Subjects with prolonged QRS duration (> 120 ms), pre-excitation syndrome, non-sinus rhythm and low-quality ECG were excluded from the analyses. After exclusions, a total of 5676 subjects remained in the cohort.

Digital 12-lead ECG was recorded using the Marquette MAC 5000 electrocardiograph (GE Marquette Medical Systems, Milwaukee, WI). Pass-band setting of 0.05-100 Hz was used. Representative beats were produced for each lead from the 10-second recording with QT Guard software (v. 1.3, GE Marquette Medical Systems). These representative leads were then used for both manual and automated grading of ER. Manual grading was performed before any algorithm development by an experienced analyst blinded to the results of the automated method. ER was graded separately for each inferior (II, III, aVF) and lateral (I, aVL, V4-V6) lead as per the recent consensus statement [5]. A subject was considered ER positive if at least two inferior or two lateral leads were positive. An example of an ER positive recording is shown in Figure 1.

2.2 Automated detection of ER patterns

The algorithm used for the automatic detection of ER is shown in Figure 2 and explained in the next sections.

2.2.1 Preprocessing of ECG

The first step is to preprocess the data by resampling the ECG recordings from 256 to 512 Hz, to improve the resolution of the QRS complex, and removing the baseline offset of the signals following the procedure [7] with an average value of the PR segment calculated in a 30 ms window. PR segment is used as the reference for the baseline offset removal to follow the electrocardiographic measures of the ER pattern published in [5]. Thereafter, R-wave detection is performed.

A rule-based filter is implemented to remove ECG recordings with low amplitude R waves and ECG signals with S waves, as they are not allowed by the ER pattern definition

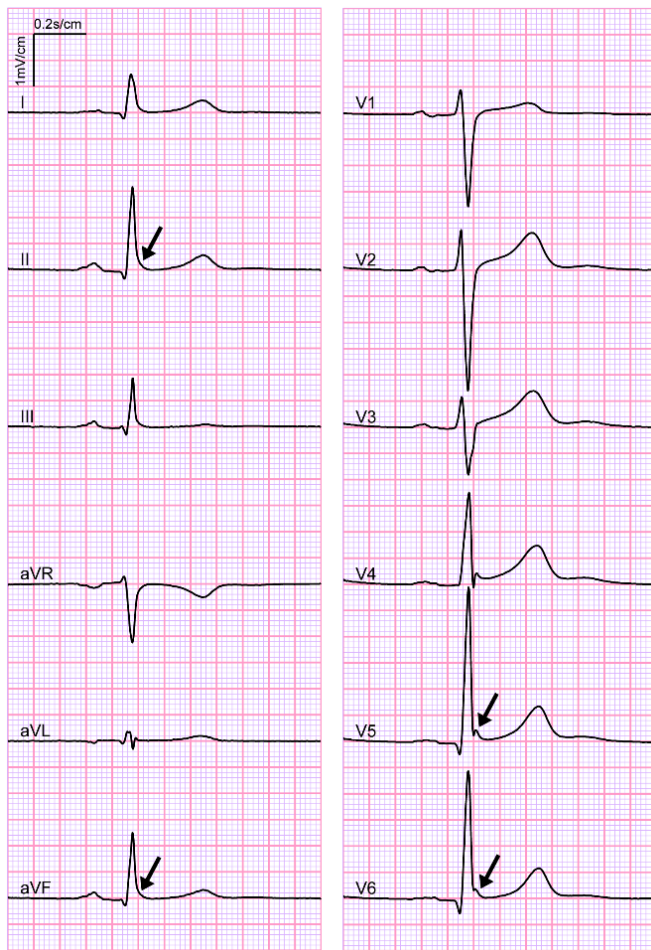


Figure 1. An example of an ECG subject with infero-lateral ER pattern. Leads II and aVF show slurred ER and leads V5 and V6 show notched pattern. Paper speed is 50 mm/s.

[5], [14]. R-peak amplitudes are used to remove ECG recordings with R-peak < 0.15 mV. Similarly, ECG signals with S waves are filtered from the data. S wave detection is performed by looking for the inflection points and zero-crossings of the ECG band-passed and differentiated signal, proposed by Pan-Tompkins [15]. The second derivative signal is used to find the inflection points. If there is a zero crossing point preceding an inflection point, within a time-window determined by the position of the R-peak, that exceeds a -0.08 mV threshold, then it is considered as an S-wave. False S waves, generated by small noise detections, are removed using the amplitude threshold level. As the removed cases cannot represent ER positive instances [5], [14], they are classified automatically as negative ER cases in our implementation.

2.2.2 Waveform prototype-based feature vector

Extraction of a waveform prototype-based feature vector consists on the acquisition of a set of prototypes that reflect the waveforms of the signals and the distribution of the data to be classified. So, each prototype should be able to represent

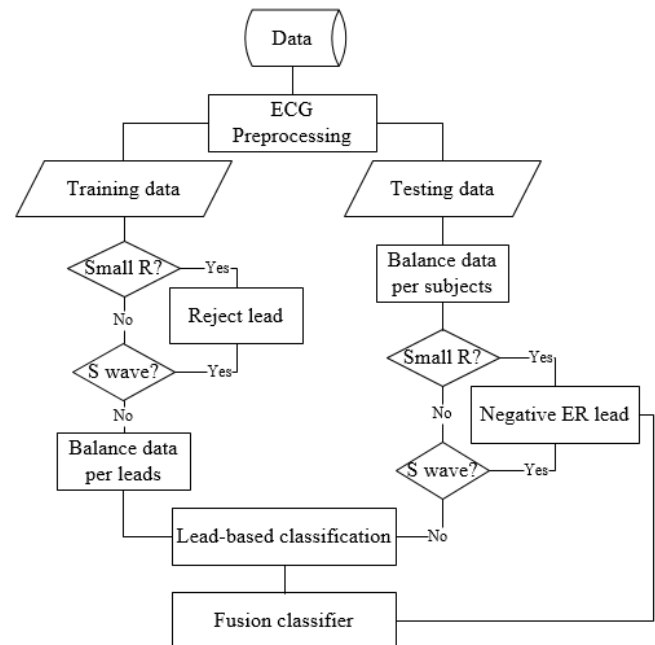


Figure 2. Schematics of the ER automatic detection algorithm.

the class it belongs to, in this case the positive and negative ER waveforms, and train the classifier to later predict the label of new data.

The prototypes are extracted per each lead independently. A prototype must not necessarily be comprised of the whole set of features describing the cases, it can be represented by only a subset of the most important features [16]. Thus, the prototype-based feature vector used is a fragment of the ECG signal where the ER pattern is located. The start point is the 50% amplitude of the R-peak in the QRS down-slope and the end point is fixed in a 25 samples window. Thereafter, the 25 samples are reduced to 15 samples using cubic spline interpolation. These 15 ECG signal samples represent the QRS-downslope changes and are used as the features for the classifier of each particular lead independently. Examples of the features extracted are shown in Figure 3 for inferior, and Figure 4 for lateral leads. Variability of the pattern among the leads related with changes in amplitude and notching or slurring configuration can be observed and represent the differences in the prototype-based feature vector extracted between normal ECG and ER leads.

2.2.3 Lead-based classification

The automatic detection of the ER pattern is first done for each lead independently using the prototype-based feature

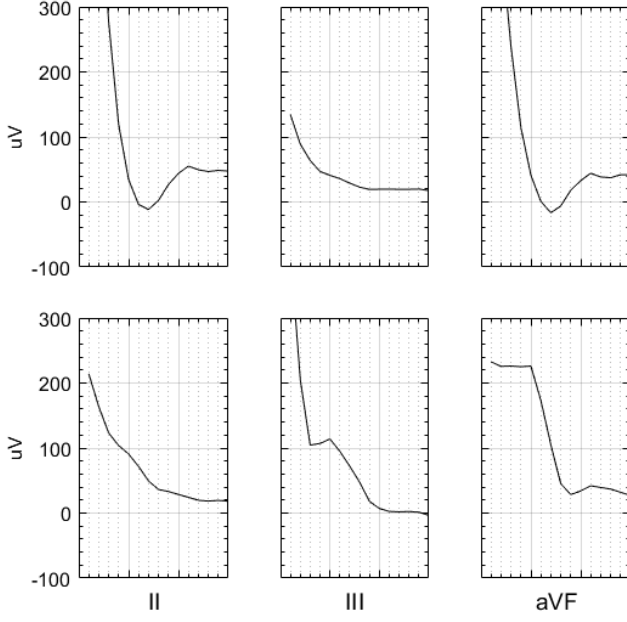


Figure 3. Example of extracted prototypes in the inferior region. Prototypes in the top row represent negative ER cases and, in the bottom, positive notch/slur ER cases.

vectors of each lead. Three types of supervised machine learning methods are compared for the classification of the eight lead waveform prototype-based feature vectors. The classifiers included are linear discriminant analysis (LDA), k-nearest neighbor (KNN) algorithm and support vector machine (SVM). KNN-classifiers are used with cityblock distance metric and 9 neighbors to predict the labels for each lead. Gaussian kernel is used in the SVM classification and linear discriminant functions are obtained to train the discriminant lead based classifiers. Performance of the classifiers is compared using Friedman test which is a non-parametric equivalent of the repeated-measures ANOVA. Friedman test ranks the algorithms separately for each dataset, according to the average error rates, by giving the smallest value (rank=1) to the one with smallest error [17].

2.2.4 Fusion classifier

Leads filtered by the rule-based filter are labeled as negative ER, because they do not represent the ER pattern, and are fused with the results of the lead-based classifiers labels predicted by the supervised classification (Figure 2). The idea of the fusion classifier is to obtain a unique decision on whether a subject has the ER pattern or not based on the results of the lead based classification. However, since the consensus definition of ER states that the detection should be done in the lateral and inferior region separately [5], the classifier fusion provides the decision for each region. Voting method is used to combine the labels of each subject. The reason to choose a voting approach for classifier fusion is based on the consensus

definition of ER with slur/notch configuration, where it is stated that at least two contiguous leads should have the pattern in order to be considered positive ER [5]. Therefore, leads are divided in inferior and lateral, and classified as positive if ≥ 2 leads in the region are predicted by the lead based classification as ER positive.

2.2.5 Classifier validation

The data is divided into testing and training parts with five folds, to perform five-fold cross validation. Additionally, the data is balanced to have the same number of instances among the two labels or classes (ER positive, ER negative). The supervised classifiers are trained with the feature vectors of the training set and the labels determined by the manual grading of ER. Then, the detection of ER is done using the feature vectors of the testing set and predicting the labels (ER positive, ER negative). For the training set, the data is balanced per each lead independently using the lead based labels after the removal of the waveforms with small R-peaks and S waves. For the testing set, the data is balanced per subjects using the final ER labels (Figure 2). The balancing of the classes is done by selecting all of the data with ER positive labels and randomly selecting the same number of ER negative cases.

2.3 Feature reduction method

Additional reduction of the number of features is considered in order to allow data visualization and reduce time complexity computation. Traditional techniques for dimensionality reduction, principal component analysis (PCA) and multidimensional scaling (MDS), are simple and efficient but can fail to find low-dimensional embedding of some data structures. Isometric feature mapping (Isomap) is a novel nonlinear manifold learning technique that preserves the intrinsic geometry of the data while reducing the dimensionality of the feature vector [18]. However, one disadvantage of using Isomap is that it does not provide the mapping to embed the original data into the lower dimensional space, which would be required for classification and visualization of new samples. This problem was solved using generalized regression neural networks (GRNN) [19] to establish the mapping.

2.4 Borderline cases assessment

SVM and LDA classifiers define a decision border between classes that it is used later for the prediction of new labels. SVM uses the support vectors to train a classifier function determine by:

$$f(x) = \sum_{i=1}^n a_i y_i G(x_i, x) + b,$$

where (a_1, \dots, a_n, b) are the estimated SVM parameters and $G(x_i, x)$ is the dot product in the Gaussian predictor space

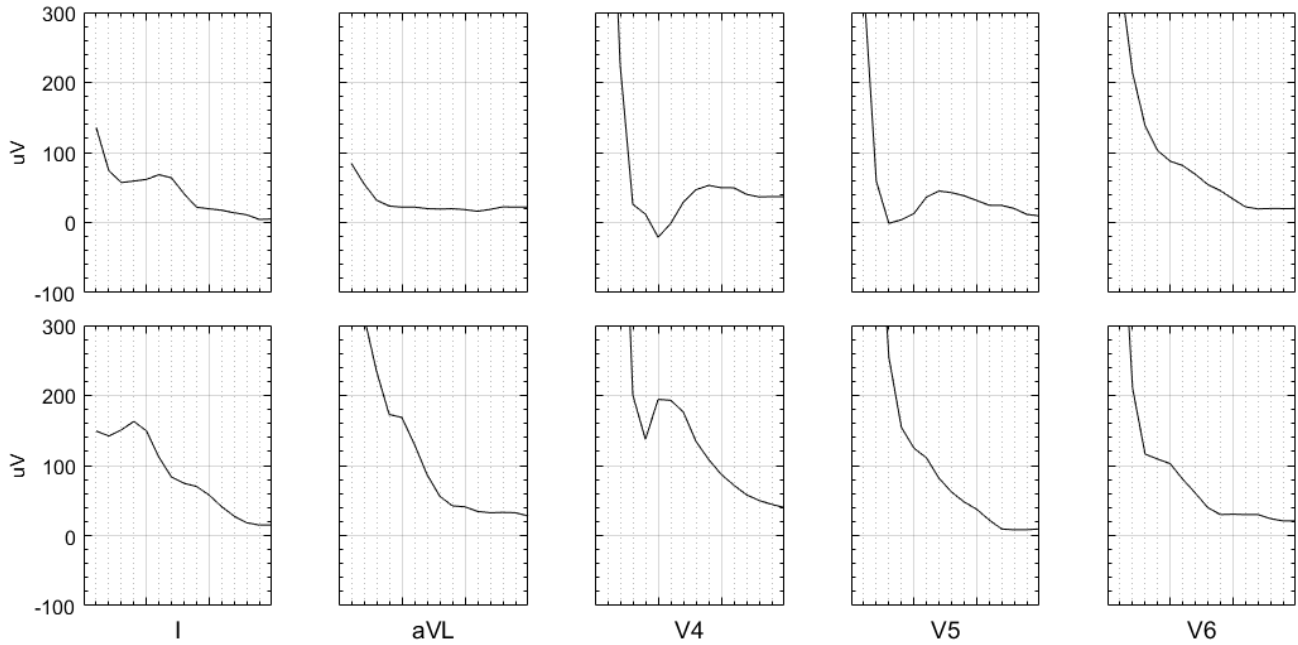


Figure 4. Example of extracted prototypes in the lateral region. Features in the top row represent negative ER cases and in the bottom positive notch/slur ER cases

between x and the support vectors. This function provides an estimate of the signed distance of x to the decision boundary which can help to analyze how far the test samples are from the SVM vectors. Similarly, we can analyze the distances of the testing data from the hyperplanes provided by the LDA classifier. The equation of the signed distance of a point to a plane in the hyperspace is:

$$D = \frac{\sum_{i=1}^n a_i x_i + c}{\sqrt{\sum_{i=1}^n a_i^2}}$$

where (a_1, \dots, a_n) represents the coefficients of the hyperplane, c is the constant and (x_1, \dots, x_n) are the coordinates of the point.

These distances to the borderline can be used to assess the confidence on the decision of the classifier and to identify borderline cases characterized by J_p amplitudes close to the threshold of 0.1 mV or slight changes in the QRS down-slope. Computation of the distance to the borderline was done for each lead independently and combined by averaging the normalized distances of contiguous leads.

3. Results

The number of subjects analyzed was 5676 representing 45408 leads. All leads were manually labelled beforehand for the supervised classification. A total of 1229 leads (2.71%) were found by the algorithm to have small R amplitudes and

S wave was detected in 23440 leads (51.62%). A total of 844 subjects were manually labeled as ER positive with 449 corresponding to inferior ER, 338 lateral and 57 cases with both. In total, 3128 leads were labeled as ER positive with the majority of the labels in leads I (518), II (681) and aVF (581), but only 41 in lead V4.

Five-fold cross validation was performed over 10 runs in order to assess the accuracies of the models by repeatedly cross-validating the classifiers following the methods described in section 2.2.5. Results for ER detection using different machine learning methods are shown in Figure 5. Range of accuracies for the classification of ER was between 87.58 and 95.93%, with mean values for inferior ER in 92.48% for KNN, 92.74% for SVM and 91.29% for LDA. Accuracies in the lateral region using KNN and SVM showed lower performance than in the inferior leads. The mean accuracies for lateral ER detection were 91.57% with KNN classification, 92.21% with SVM and 91.48% with LDA.

Sensitivity, defined as the number of ER cases detected by the algorithm over the total number of positives; and specificity, defined as the number of non ER cases detected over the total number of ER negatives, are shown in Figure 6 and Figure 7,

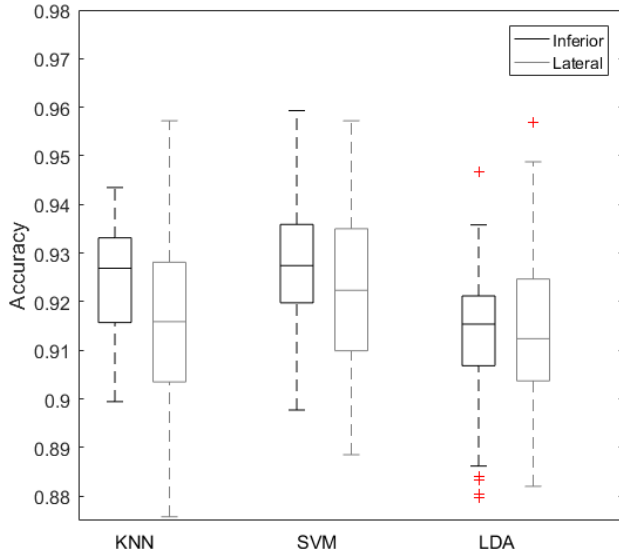


Figure 5. Classification accuracy of ER using the KNN, SVM and LDA algorithms over fivefold cross-validation with 10 repetitions. Each boxplot indicates the median (50th percentile) and interquartile range (25th and 75th percentile) with outliers marked with +.

respectively. Highest sensitivity for all of the classifiers was 100% and lowest was 78.84% in the lateral ER detection using LDA. Mean values of sensitivity in inferior ER were 93.75% using KNN, 91.87% using SVM and 88.31% with LDA classification. Highest mean sensitivity in the ER lateral region was accomplished using KNN algorithm (93.36%) and lowest using SVM (91.74%). Specificity in the inferior leads showed better results than in the lateral territory. The range of specificity was between 86.94 and 96.15%. Best average specificity was accomplished in the inferior leads (93.13%) and in lateral region (92.34%) using SVM. Lowest average specificity was obtained using KNN in the ER inferior detection (91.94%) and in the lateral (91.01%).

Statistical significant ($p < .05$) differences were found between the accuracies of the classifiers using the Friedman test. Thus, we performed Bergman-Hommel's post-hoc test, since it is the most powerful one, and recommended when the differences among the classifiers compared are not substantial and the number of classifiers is not very large [20]. Table 1 shows the results of the Bergman-Hommel's post-hoc test with ones indicating significant differences and zeros otherwise; ranking SVM as the best performing algorithm followed by KNN and LDA.

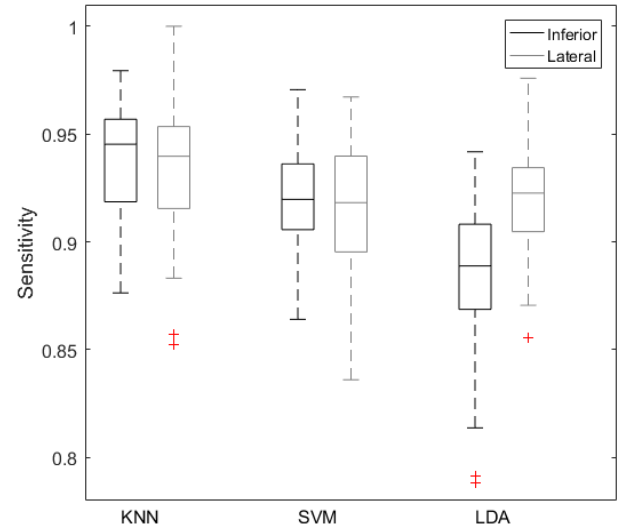


Figure 6. Classification sensitivity of ER using the KNN, SVM and LDA algorithms over fivefold cross-validation with 10 repetitions. Each boxplot indicates the median (50th percentile) and interquartile range (25th and 75th percentile) with outliers marked with +.

3.1 Comparison with automatic method based on the analysis and quantification of the notch or slur slope

The method presented by Kenttä et al. [7] based on the analysis and quantification of the notch or slur slope, was tested with the data used in this paper in order to compare its

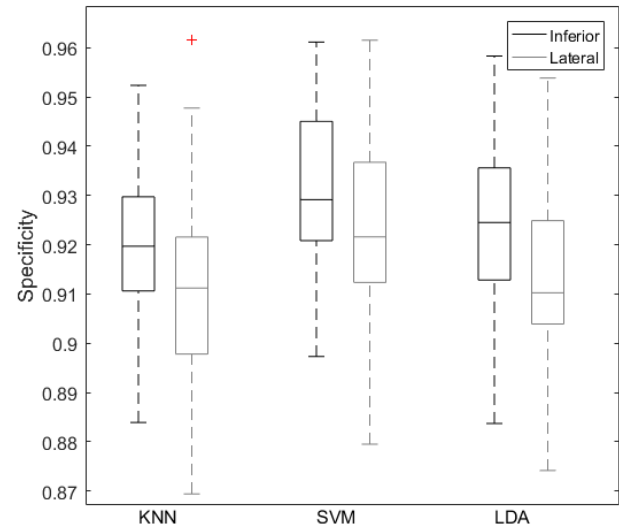


Figure 7. Classification specificity of ER using the KNN, SVM and LDA algorithms over fivefold cross-validation with 10 repetitions. Each boxplot indicates the median (50th percentile) and interquartile range (25th and 75th percentile) with outliers marked with +.

Table 1. Results of Bergman-Hommel's post-hoc test to indicate whether there are statistical significant differences (ones) between the classifiers or not (zeros)

Classifier	KNN	SVM	LDA
KNN	0	1	1
SVM	1	0	1
LDA	1	1	0
Average ranks	1.7	1.3	3.00
Final rank	2	1	3

performance. Their algorithm yielded a sensitivity of 63.83% and a specificity of 95.67%, with values for inferior ER of 66.80% and 97.35%, and for lateral ER of 57.72 and 98.02%, respectively. Our average performance using SVM showed results of 91.87% and 91.74% for inferior and lateral sensitivity respectively. Average specificity using SVM was 93.13% in inferior leads and 92.34% in lateral leads. Even though the method by Kenttä et al. showed highest specificity, the detection ability of ER is determined by the sensitivity of the algorithm, which reflects the ratio between the true positives detected and the total positives.

3.2 Reduction of the features

Three dimensional representation of the extracted features was selected based on the residual variance reduction using the Isomap algorithm on our data (Figure 8). Mapping of the new data was accomplished using GRNN. We constructed the GRNN in each validation iteration with the 15D prototype-based input vectors and the 3D output vectors from Isomap algorithm of each lead. Then, the networks were used to map the lead-based testing data to the lower dimensional space to generate reduced-dimensional prototypes for assessing classification performance. Additionally, GRNN were used to establish the mapping in the inverse direction, i.e. back from 3D into 15D feature space, followed by classification experiments with the reconstructed prototypes, in order to test how efficiently the nonlinear Isomap maintains the discriminating characteristics of the original data. GRNN were constructed with the same training data as above. Performance of the classifiers, using the 3D prototypes and the reconstructed 15D prototypes, are shown in Table 2.

The ER detection performance was comparable to the one with the original 15D prototype data, indicating that the Isomap maintains critical information in the waveforms, and the detection algorithm could be implemented in 3D feature space without a compromise in accuracy while gaining computational complexity in classification.

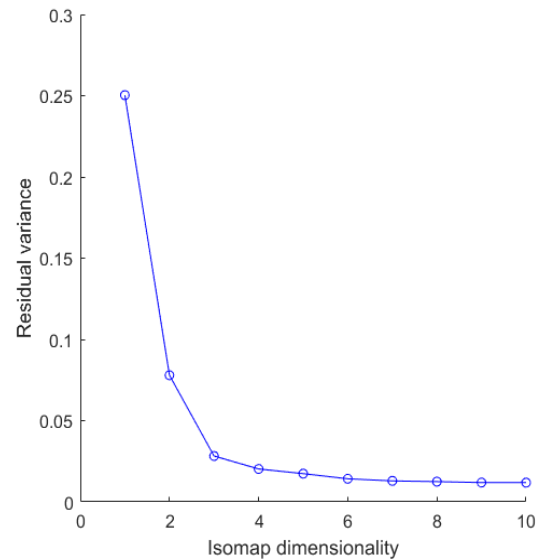


Figure 8. Residual variance of the Isomap dimensionality embedding data.

Visualization of the 3D embedding using Isomap for lead I is shown in Figure 9. Linear function determined by the LDA classifier shows the separation boundary between the two classes but it can be observed that the two classes are overlapping, leading to errors in the classification.

3.3 Analysis of borderline cases

Analysis of critical cases was performed according to the distances of the prototypes to the classifier decision boundary. Critical cases, defined as the ones with distances close to the boundary, are difficult to classify by the algorithm and even by medical professionals.

Table 2. Sensitivity (Sn) and Specificity (Sp) of the three different classifiers with the 3-dimensional prototypes and the 15-dimensional reconstructed prototypes.

Classifier	Inferior		Lateral	
	Sn (%)	Sp (%)	Sn (%)	Sp (%)
KNN 3D	92.76	89.40	93.08	86.82
SVM 3D	93.10	90.05	92.88	87.56
LDA 3D	89.05	91.08	91.99	87.43
KNN 15D Reconstructed	93.18	89.14	93.50	86.59
SVM 15D Reconstructed	93.73	89.60	93.71	87.04
LDA 15D Reconstructed	88.81	91.08	92.38	87.28

Figure 10 shows the histogram of the distances to the SVM classifier decision boundary, and it can be seen that incorrectly classified subjects are closer to the border. Thus, manual verification of the ER pattern (or using extra clinical data available) could be performed by the doctor in cases advised by the classifier to decide whether or not the subject has the pattern.

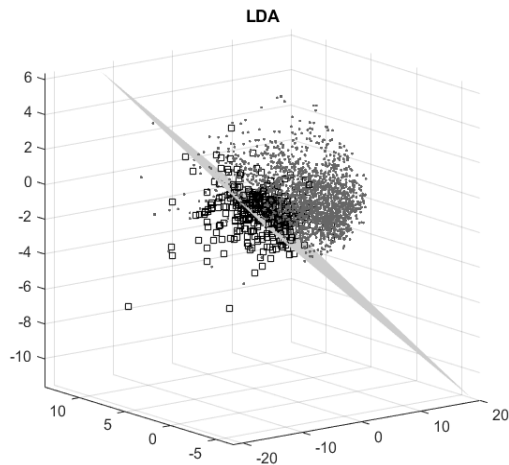


Figure 9. Visualization of the embedding feature vector for Lead I. Lighter grey points (.) represent the ER negative cases and black squares (□) the positive ER. Plane separating the two classes represents the linear discriminant function.

For example, if we extract the cases in the inferior region with distances between 0 and 0.2, 85 subjects from 1692 in total are selected for manual revision, from where 20 were false negatives detected by the algorithm and 7 were false positives. By combining the classifier output and the doctor's manual verification in the indicated critical cases, the clinical decision making accuracy could improve much. In fact, the sensitivity of the detection could increase up to 95.06% and the specificity up to 94.77% in the inferior territory. Figure 11 shows a critical inferior case selected by the algorithm. It can be observed that the amplitude of Jp is close to the 100 μ V threshold, which makes it difficult to label even manually. In the lateral region, it is more challenging to find the incorrectly misclassified cases since the number of leads is bigger and the presence of S waves is more common so we cannot include them in the analysis of the distances. From 91 cases with distances between 0 and 0.2 in the lateral leads, 3 were false negatives identified by the algorithm and 25 false positives, which could lead to an improvement on the sensitivity of up to 0.76% and on the specificity of up to 1.93%. Similar results were obtained using the distances to the hyperplane constructed by LDA.

4. Discussion

The definition of the ER pattern has been a controversy over the last decade. The term ER was first used to describe ST-

segment elevation without chest pain, but it was used later to refer to terminal QRS slurring or notching without the presence of ST-segment elevation. In 2016, the American Heart Association (AHA) made a scientific statement to address the confusion and proposed to include those two different phenomenon in the ER term [14].

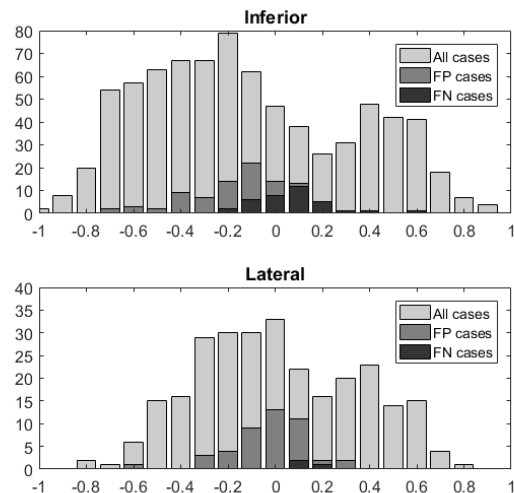


Figure 10. Histogram of the signed averaged normalized distances to the SVM classifier decision boundary. Cases with S wave and R peaks <0.15 mV are not included.

Currently, commercial ECG machines automatically detect nonspecific ST-segment elevation as ER but not the terminal QRS slurring or notching [14]. Automatic methods proposed to detect the terminal notch/slur pattern stated the need for an agreed definition and a threshold level for the peak of the notch and onset of the slur [6], [7]. The consensus definition of ER with notch/slur published in [5] and the AHA statement [14], agreed that there should be an elevation of the Jp of ≥ 1 mm above the isoelectric line to be considered as ER pattern. This is the first automated detection method after the publication of the consensus definition of the ER with slurring and notching configuration, using manual labeled data, with a Jp threshold of 0.1 mV, to train a supervised classifier.

The classifiers were trained with waveform prototype-based features composed of fragments of the ECG signals. The advantage of using these prototypes, for the pattern recognition system, is that there is no need for feature extraction based on measurements of the signals. The automatic detection is a memory-based approach, which demonstrates that the ECG signal fragment itself can be used directly to train a classifier and discriminate between classes. By using prototypes it is no longer required to measure directly the amplitude of Jp and compare it with thresholds. However, when acquiring the features from the signal, one of the critical steps is to remove the baseline offset of the signals

so that the amplitudes of the prototypes match and the comparison can be performed.

The proposed method trains and creates a model for each lead independently considering the slurred morphologies changes between the inferior leads and precordial leads, which might be related to the differences in signal amplitudes [7].

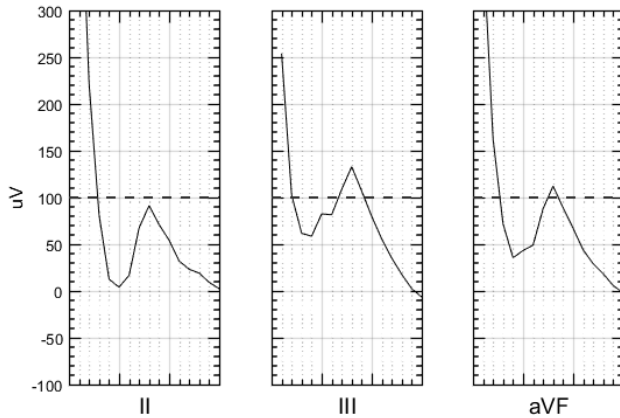


Figure 11. Example of an inferior critical case found by the algorithm based on the distance to the classifier decision boundary.

Compared with the automated detection methods based on the analysis and quantification of the signal waveforms and time measurements, machine learning approaches provide a more general way to include different patterns and variations, and it can be modified afterwards to include new morphologies in the training set. Binary classification was performed but other classes could have been involved, for example to distinguish between slur and notching configuration of the ER or to include ST-segment elevation.

Detection of lateral ER was slightly less accurate than inferior ER. However, the latter is the most significant ER pattern because it indicates the greatest risk for arrhythmic events [7]. Overall accuracies of the different classifiers showed results around 90%, indicating that the prototype-based feature vector is a good representation of the differences between ECG signals with and without the ER pattern. Friedman test found statistical significant differences between the accuracy of the classifiers, showing that SVM was the one with best performance. Additionally, an advantage of SVM over KNN is the measure of the distances to the borderline, in order to assess the critical cases. The measurement of the distances to the classification borderline was found to be an effective way to further review critical cases and identify false negative detections, especially in the inferior region.

Lower dimension representation of the prototype-based feature vector was accomplished using Isomap and GRNN, without affecting the performance of the automatic detection

of the ER pattern. Thus, nonlinear mapping is an effective way of dimensionality reduction of the features and can be used for clustering visualization and reduce time complexity computation of the classification. However, when mapping the original features to the 3D, the physiological interpretation and representation of the ECG fragment may be lost. In order to verify if critical waveform information is maintained during Isomap transform, reconstruction of the 15D prototypes was accomplished using GRNN and tested using the three different classifiers. Performance of the ER detection showed similar results using the reconstructed features and the embedding data (Table 2).

Misclassifications were mainly caused by corrupted signals, false slurred detections and borderline cases of Jp threshold. Averaging ECG complexes could induce random noise, especially when there is powerline interference, muscle noise or poor alignment of the signals, making the identification of ER morphologies more difficult. Slurred patterns are more complex to detect than notched terminal QRS since they have a wide range of variations [7]. Additionally, manual labelling by only one experienced reader can contribute to human errors in the manual detection of the ER especially in cases close to the Jp threshold. Manual labeling of ER is more difficult in the lateral region since there is more variation of the signal amplitude and represent a bigger number of leads. Identification of critical cases based on the distances to the borderline contributes to identify the subjects that are difficult to label and require further evaluation by medical professionals according to the recommendations [14].

5. Conclusion

The ER detection method proposed here uses a waveform prototype-based feature vector extracted from the ECG signal fragment where the ER is located. The method provides identification of critical cases based on the distances to the classifier decision boundary which further improves the sensitivity and specificity of the detection. Sensitivity values reaching 92-95% indicate that the proposed method could be used as an effective prescreening tool of ER.

Acknowledgements

Business Finland is gratefully acknowledged for the financial support in the VitalSens project.

References

- [1] M. Haïssaguerre, N. Derval, F. Sacher, L. Jesel, I. Deisenhofer, L. De Roy, P. Mabo, S. Matsuo, V. Probst, D. Ph, S. Le Scouarnec, D. Ph, P. Defaye, J. Schlaepfer, T. Rostock, D. Lacroix, D. Lamaison, T. Lavergne, Y. Aizawa, A. Englund, F. Anselme, M. O. Neill, M. Hocini, G. D. Veenhuyzen, P. Bordachar, M. Chauvin, P. Jais, G. Coureau, D. Ph, G. Chene, D. Ph, G. J. Klein, and J. Clémenty, "Sudden Cardiac Arrest Associated with Early Repolarization," *N. Engl. J. Med.*, vol. 358, pp. 2016–2023, 2008.

- [2] M. J. Junttila, J. T. Tikkanen, T. Kenttä, O. Anttonen, A. L. Aro, K. Porthan, T. Kerola, H. A. Rissanen, P. Knekt, and H. V. Huikuri, "Early repolarization as a predictor of arrhythmic and nonarrhythmic cardiac events in middle-aged subjects," *Hear. Rhythm*, vol. 11, no. 10, pp. 1701–1706, 2014.
- [3] J. T. Tikkanen, O. Anttonen, M. J. Junttila, A. L. Aro, T. Kerola, H. A. Rissanen, A. Reunanen, and H. V. Huikuri, "Long-Term Outcome Associated with Early Repolarization on Electrocardiography," *N. Engl. J. Med.*, vol. 361, pp. 2529–2537, 2009.
- [4] Y.-J. Cheng, X.-X. Lin, C.-C. Ji, X.-M. Chen, L.-J. Liu, K. Tang, and S.-H. Wu, "Role of Early Repolarization Pattern in Increasing Risk of Death," *J. Am. Heart Assoc.*, 2016.
- [5] P. W. Macfarlane, C. Antzelevitch, M. Haissaguerre, H. V. Huikuri, M. Potse, R. Rosso, F. Sacher, J. T. Tikkanen, H. Wellens, and G. Yan, "The Early Repolarization Pattern," *J. Am. Coll. Cardiol.*, vol. 66, no. 4, pp. 470–477, 2015.
- [6] E. N. Clark, I. Katibi, and P. W. Macfarlane, "Automatic detection of end QRS notching or slurring," *J. Electrocardiol.*, vol. 47, pp. 151–154, 2014.
- [7] T. Kenttä, K. Porthan, J. T. Tikkanen, H. Väänänen, L. Oikarinen, M. Viitasalo, H. Karanko, M. Laaksonen, and H. V. Huikuri, "Sensitivity and Specificity of Automated Detection of Early Repolarization in Standard 12-Lead Electrocardiography," *Ann. Noninvasive Electrocardiol.*, vol. 20, no. 4, pp. 355–361, 2015.
- [8] P. Aagaard, E. Shulman, L. Di Biase, J. D. Fisher, J. N. Gross, F. Kargoli, S. G. Kim, E. C. Palma, K. J. Ferrick, and A. Krumer, "Prognostic Value of Automatically Detected Early Repolarization," *Am. J. Cardiol.*, vol. 114, pp. 1431–1436, 2014.
- [9] M. Sansone, R. Fusco, A. Pepino, and C. Sansone, "Electrocardiogram Pattern Recognition and Analysis Based on Artificial Neural Networks and Support Vector Machines: A Review," *J. Healthc. Eng.*, vol. 4, no. 4, pp. 465–504, 2013.
- [10] M. Ma, M. Shao, X. Zhao, and Y. Fu, "Prototype Based Feature Learning for Face Image Set Classification," *10th IEEE Int. Conf. Work. Autom. Face Gesture Recognit.*, pp. 1–6, 2013.
- [11] B. U. Köhler, C. Hennig, and R. Orglmeister, "The principles of software QRS detection," *IEEE Eng. Med. Biol. Mag.*, vol. 21, no. 1, pp. 42–57, 2002.
- [12] A. Aromaa and S. Koskinen, *Health and functional capacity in Finland: Baseline Results of the Health 2000 Health Examination Survey*. Helsinki: National Institute for Health and Welfare, 2004.
- [13] S. Heistaro, *Methodology Report: Health 2000 Survey*. Helsinki: Publications of the National Health Institute, 2008.
- [14] K. K. Patton, P. T. Ellinor, M. Ezekowitz, P. Kowey, S. A. Lubitz, M. Perez, J. Piccini, M. Turakhia, P. Wang, and S. Viskin, "Electrocardiographic Early Repolarization," *Circulation*, vol. 133, pp. 1520–1529, 2016.
- [15] J. Pan and W. J. Tompkins, "A Real-Time QRS Detection Algorithm," *IEEE Trans. Biomed. Eng.*, vol. BME-32, no. 3, pp. 230–236, 1985.
- [16] P. Perner, "Prototype-based classification," *Appl Intell*, vol. 28, no. June 2007, pp. 238–246, 2008.
- [17] J. Demsar, "Statistical Comparisons of Classifiers over Multiple Data Sets," *J. Mach. Learn. Res.*, vol. 7, pp. 1–30, 2006.
- [18] J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A Global Geometric Framework for Nonlinear Dimensionality Reduction," *Science (80-.)*, vol. 290, pp. 2319–2323, 2000.
- [19] J. Kortelainen, E. Väyrynen, and T. Seppänen, "Isomap Approach to EEG-Based Assessment of Neurophysiological Changes During Anesthesia," *IEEE Trans. NEURAL Syst. Rehabil. Eng.*, vol. 19, no. 2, pp. 113–120, 2011.
- [20] S. Garcia and F. Herrera, "An Extension on 'Statistical Comparisons of Classifiers over Multiple Data Sets' for all Pairwise Comparisons," *J. Mach. Learn. Res.*, vol. 9, pp. 2677–2694, 2008.